Global Jour. of Engg. & Tech.
Volume 2, Number 2 (2009) 211-221

AN ARTIFICIAL NEURAL NETWORK MODEL FOR FORECASTING DAILY GLOBAL SOLAR RADIATION IN IBADAN, NIGERIA

D. A. FADARE and T. T. OLUGASA
Department of Mechanical Engineering, University of Ibadan, Ibadan, Nigeria

Abstract : Solar radiation, the primary driver for many physical, chemical and biological processes on the earth's surface, is considered the most indispensable parameter in the performance prediction of solar power systems. In this study, an artificial neural network (ANN) model was developed for predicting missing solar radiation data for Ibadan (Lat. 7.43°N; Long. 3.9°E; Alt. 227.2m), Nigeria. This study utilized daily solar radiation data for the period of 1984 to 2007 (24 years) from a meteorological station in Ibadan. The ANN model was designed using the Matlab® Neural Network Toolbox and five different structures of the model were investigated. Structure 1 utilized solar radiation data for 5 days to predict the next 25 days expected data; structure 2 utilized data for 10 days to predict the next 20 days; structure 3 used data for 15 days to predict succeeding 15 days; structure 4 used data for 25 days to predict next 5 days data; structure 5 used data for 5 days to predict the next 1 day solar radiation. The different structures were trained by using solar radiation data for 22 years and one year and the prediction accuracies were evaluated using the solar radiation values for year 2007. Results showed that structure 5 with correlation coefficient of 0.73 and 0.79 when trained with 22 years and 1 year, respectively gave the best prediction performance. Thus, indicating the suitability of structure 5 for prediction of solar radiation missing data.

1 INTRODUCTION

Modelling and prediction of daily global solar radiation data are essential features in the development and assessment of solar powered systems, such as solar thermal and solar PV systems with applications in meteorology, renewable energy and solar conversion energy (Guofeng et al., 2007). The need for comprehensive and accurate solar radiation data on long term for any given cite is considered one of the most important meteorological parameter in the performance prediction of renewable energy systems, particularly in sizing photovoltaic (PV) power systems (Mellit et al., 2006). However, missing data is commonplace problem in meteorological records, which causes distortion in data analysis. It is fairly common for a time series to have gaps for a variety of reasons. Usual reasons include the absence of solar data at the needed frequency, registration errors, and as a result of the deletion of outliers (Little and Rubin, 1987; Shafer, 1997; Greene, 2000). For instance, the daily global solar radiation data collected from a meteorological station located in Ibadan, Nigeria for the period between 1984 and 2007 had 5.09% missing data with the data for year 2005 completely missing from the database.

A possible solution to this problem is the interpolation and extrapolation method in which prediction estimates are made using data from days before and days after the missing data (interpolation) or using previous data before the missing data (extrapolation). In this method, a

Key words : Solar radiation, artificial neural network, forecasting, renewable energy, Ibadan.

univariate and linear dependence in the dataset is assumed (Pourhamadi, 1989; Kohn and Ansley, 1986). However, meteorological data are known to be multivariate and non-linear (Acock and Pachepsky, 1999), thus limiting the accuracy and reliability of the estimates. A second approach is the use of statistical models, such as Auto-regressive (AR) prediction (Aguiar and Collares-Pereira, 1992), Markov Chains (Maafi and Adane, 1989), Auto-regressive moving average (ARMA) models, (Mora and Sidrash-de-Cardona, 1998), and Markov transitions matrix (MTM), (Aguiar et al., 1988). All these developed models were based on simplified statistical assumptions, which in most cases are not always true (Guessom et al., 1998).

Adaptation of data from neighbouring weather stations is another method used in reconstructing missing data (Mott et al., 1994). Ashraf et al. (1997) used geostatistics to inter-relate data from several neighbouring weather stations. When the density of the weather stations is low, the use of data from neighbouring stations becomes problematic because daily weather patterns are often strongly related to local landscape conditions (Blackie and Simpson 1993). The "outlier"-handling technique is another method used for predicting missing data. This method was developed only for the univariate time series. But the forecasting performance of the model was not evaluated (Stockinger and Dutter, 1987). Double vector quantization forecasting method based on Kohonen self-organizing maps (Simon, 2004) and the use of linear and mixed models are other methods used. All these models are for short term forecasting. Hence, are not applicable for prediction of long term missing solar radiation data.

The artificial neural network (ANN) approach provides a viable solution to the problem of prediction of long term missing data because it is based on training not on statistical assumptions. ANN model can be trained to predict results from examples and once trained can perform predictions at very high speed (Mellit et al., 2005). They are able to deal with non-linear problems such as multivariate time series prediction of solar radiation. A number of different architectures and learning methods have been applied with varying degrees of success to the problem of predicting future values of time series for solar radiation. (Elizondo et al., 1996; Al-Alawi and Al-Hinai, 1998; Mellit, 2006; Fadare, 2009).

ANN often called a "neural network" (NN), is a mathematical model or computational model based on biological neural networks. It is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like humans, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well (Stergiou and Siganos, 1996)

ANN is an adaptive system that changes its structure, based on external or internal information that flows through the network during the learning phase. Adaptive means that the system parameters (weight and bias) are changed during operation called the training phase. The network can be trained through supervised or unsupervised learning algorithm. After the training phase, the ANN parameters are fixed and the system is deployed to solve the problem at hand (the testing phase). The ANN system parameters are changed systematic in a step by step procedure to optimize the performance criterion or to follow some implicit internal constraint, which is commonly referred to as the learning rule. The input/output training dataset are so fundamental in neural network technology, because they convey the necessary information to identify the complex relationship between the input parameters and the desired output parameters. The nonlinear nature of the ANN processing elements provides the system with lots of flexibility to achieve practically any desired input/output (http://www.learnartificialneuralnetworks.com/).

The essence of this study was to investigate the feasibility of using ANN to model the non-linear relationship in the time series global solar radiation. Hence, the model can be used to predict the missing daily global solar radiation data in Ibadan, Nigeria.

2 Materials and Methods

2.1 Data Source

The daily global solar radiation data used in this study was obtained from a meteorological station located in Ibadan (Lat. 7.43°N; Long. 3.9°E; Alt. 227.2m), Nigeria, for the period of 1984 to 2007 (24 years). About 5.09% of the solar radiation data were missing for the period. Statistics of the missing data for each year is shown in Table-1.

Table-1 : **Statistics of the missing solar radiation data in each year**

S/N	Year	No. of missing data	% of missing data	S/N	Year	No. of missing data	% of missing data
1.	1984	4	1.09	13.	1996	0	0
2.	1985	5	1.37	14.	1997	0	0
3.	1986	22	6.03	15.	1998	2	0.55
4.	1987	4	1.10	16.	1999	0	0
5.	1988	4	1.09	17.	2000	0	0
6.	1989	10	2.74	18.	2001	0	0
7.	1990	3	0.82	19.	2002	0	0
8.	1991	12	3.29	20.	2003	0	0
9.	1992	3	0.82	21.	2004	0	0
10.	1993	10	2.74	22.	2005	365	100
11.	1994	3	0.82	23.	2006	0	0
12.	1995	0	0	24.	2007	0	0

It can be seen that 13 years out of 24 years have missing data, with the data for year 2005 completely missing. The overall percentage of missing data for the period 1984–2007 was estimated as 5.09%. The total number of 8,289 available solar radiation values was used in this study. Figure 1 shows the variation of daily global solar radiation for ten different years with complete solar radiation data.

2.2 Design of the ANN model

Neural Network Toolbox for MATLAB® was used to design the neural network.

The basic steps involved in designing the network were : Collection of solar radiation data; Pre-processing of data (removal of missing data); Design of the neural network; Training and testing of the neural network; Simulation and prediction with the neural networks and Analysis and post-processing of predicted result.

2.2.1 Pre-processing of data

The solar radiation data collected had 5.09% missing data. Before the data can be used to train the network, the missing data must be expunged. This was done by transferring the data to excel worksheets and deleting the days with missing data, leaving complete data with no gaps. The data was then partitioned in to training and test datasets.

2.2.2 Design of the neural network

The commonest type of artificial neural network consists of three layers of neurons : a layer of *"input"* neurons is connected to a layer of *"hidden"* neurons, which is connected to a layer of *"output"* neurons. The input neurons represent the raw information that is fed into the network. The activity of each hidden neuron is determined by the activities of the input neurons and the weights on the connections between the input and the hidden neurons. The behaviour of the output neurons depends on the activity of the hidden neurons and the weights between the hidden and output neurons. Figure-2 shows the architecture of a typical three layered ANN model.

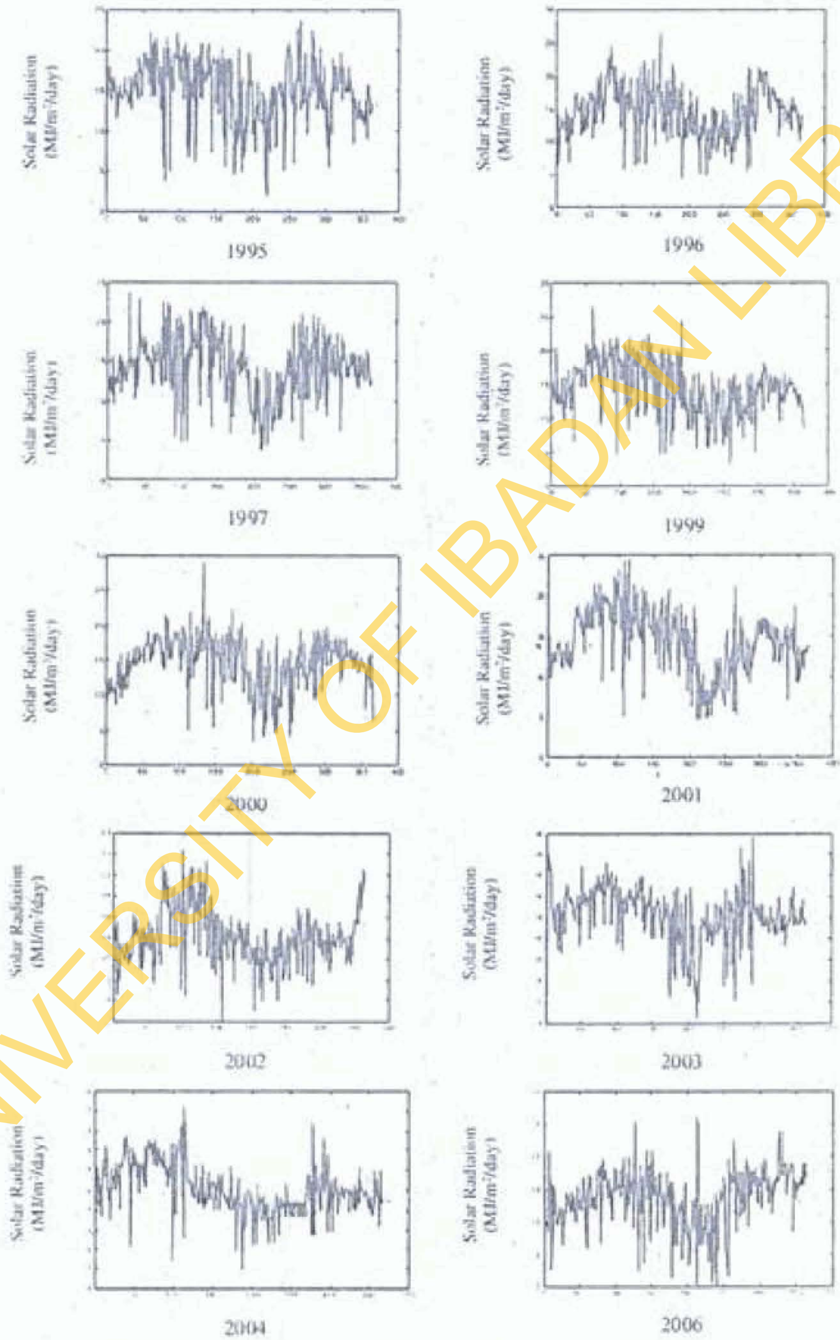

Figure-1 : Variation in daily solar radiation at Ibadan for different years with no missing data

The number of hidden layers and the number of neurons in the hidden layer was varied during the learning phase so that the network can be trained efficiently. The hidden layer was varied from 5 to 20 in steps of 5. A constructed neural network must be trained to learn or model the relationship between given input/output data or set of measurements. A portion of the available data was used as the training dataset, while a smaller portion of the data that was excluded from the training dataset was used as the test dataset. In this study, two datasets of daily global solar radiation data were used to train the network. The first set consisted of data for 22 years spanning 1984–2006 (excluding 2005 data) and the second dataset consisted of data for one year (2001).

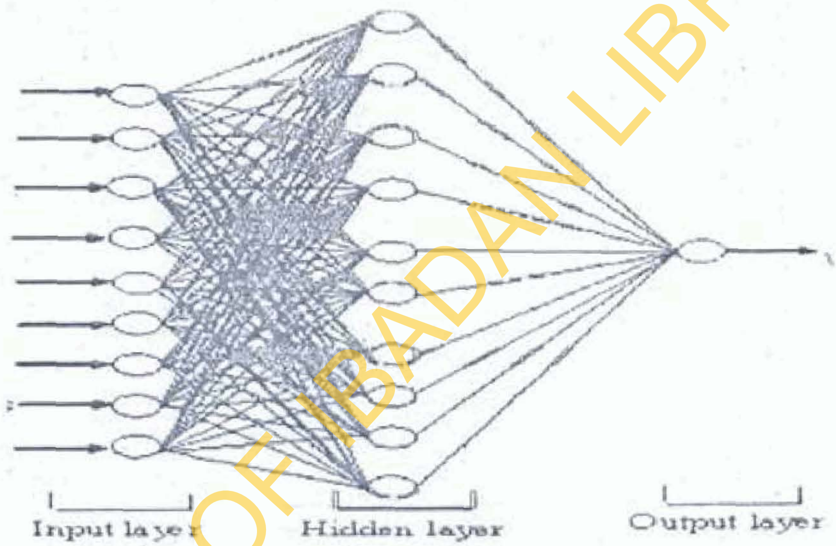

Figure-2 : Network architecture for a multi layer ANN

The two training datasets have been used to train the network according to following structures :

Structure 1 : The first 5 values (first 5 days) of solar radiation data from each month was used in the input layer and solar radiation values the remaining 25 days for each month was in the output layer for all months of the year.

Structure 2 : The first 10 values of solar radiation data from each month was used in the input layer and the rest of the values for the remaining 20 days was used in the output layer for all months of the year.

Structure 3 : 15 values of solar radiation data was used in the input layer and the values of the remaining 15 days was used in the output layer for all months of the year.

Structure 4 : The first 25 values of solar radiation from each month was used in the input layer and the solar radiation values for the remaining 5 days was used in the output layer for all months of the year.

Structure 5 : 5 values of solar radiation were used in the input layer and in the output layer. the next solar radiation value.

The performance of the different networks in predicting the daily global solar radiation for the year 2007 was tested to determine the network's generalization capability. The early stopping technique was used to improve generalization capability of the network and to prevent the network from overfitting the training dataset. Overfitting occurs when the network becomes too used to the training data set and thus fails in predicting other data outside the training dataset. (Demuth and Beale, 2000).

2.2.3 Training of the neural network

The network was trained by feeding in some teaching patterns and letting it change its weights according to some learning rule (http://www.learnartificialneuralnetworks.com/). The Levenberg-Marquardt *(trainlm)* training algorithm was used because in comparison with other algorithms like BFGS Quasi Newton, Resilient Back-propagation, Scaled conjugate gradient etc., it converges faster and in most cases, it has the best performance. Though it has the disadvantage of using up a lot of memory and if the network is large, then it runs out of memory. (Demuth and Beale, 2000).

For the training of each structure, the tan-sigmoid transfer function *'tansig'* was used in the hidden layer, while linear transfer function *'purelin'* was used in the output layer. The *'purelin'* transfer function was used so that the output would not be limited like the *'tansig'* function which generates output between 0 and +1. If linear output neurons are used, the output can take on any value. (Demuth and Beale, 2000).

2.2.4 Testing of the ANN model

The mean square error (MSE) between the network predicted outputs and the desired outputs was used as the performance function during the training phase. The training was terminated when the threshold of MSE = 0.001 or when the number of iterations is equal to 1000 is attained. The performance of the networks with different structures and varying number of neurons in the hidden layer trained with the 22 years and 1 year solar radiation data were tested with the correlation coefficient between the predicted and the measured values in prediction for 2007 daily global solar radiation.

3 RESULTS AND DISCUSSIONS

The performance parameters of the different network structures trained with the 22 years solar radiation data for the prediction of 2007 data are presented in Table-2. The table shows the number of iterations during the training phase, and the correlation coefficient (R-value) and mean square error (MSE) between the ANN predicted and actual solar radiation values. In the table, structure 5 network with 5 neurons in the input layer, 5 neurons in the hidden layer, and 25 neurons in the output layer was designated by 5-5-25. It can be seen (Table 2) that the number of iterations ranged from 10 - 19, 8 - 13, 9 - 13, 10 17 and 9 -12, respectively for networks with structure 1, 2, 3, 4 and 5. Thus, indicating the fast convergence of structure 5 compared with other structures. The R-value for the networks ranged from 0.63 – 0.72, 0.43 - 0.71, 0.32 – 0.75, 0.50 – 0.68 and 0.66 – 0.73, respectively for structure 1, 2, 3, 4 and 5, while the corresponding MSE were ranged between 17.70 – 22.17, 14.40 – 21.09, 12.19 – 20.00, 9.33 – 11.06 and 4.33 – 5.79, respectively. The results showed that structure 5 had higher R-values (0.66 – 0.73) and lower MSE values (4.33 – 5.79), thus indicating structure 5 as the best network compared to other structures. Figure-3 shows the comparison between the actual and ANN predicted daily solar radiation for the year 2007 using the network with structure 5 (with 5-10-1 architecture), while Figure 4 shows the daily variation in the predicted and the actual values.

Table-2 : **Performance parameters of the different network structures trained with 22 years solar radiation data for prediction of 2007 data**

Structure	No of iteration	Correlation Coefficient (R-value)	MSE
Structure 1			
5-5-25	19	0.72	19.12
5-10-25	11	0.72	17.70
5-15-25	14	0.68	19.57
5-20-25	10	0.63	22.17
Structure 2			
10-5-20	13	0.71	14.40
10-10-20	10	0.63	19.04
10-15-20	11	0.61	16.35
10-20-20	8	0.43	21.09
Structure 3			
15-5-15	10	0.74	12.19
15-10-15	9	0.69	12.80
15-15-15	13	0.32	20.00
15-20-15	9	0.57	18.31
Structure 4			
25-5-5	10	0.68	11.06
25-10-5	17	0.67	9.33
25-15-5	17	0.50	11.41
25-20-5	12	0.68	7.64
Structure 5			
5-5-1	12	0.66	4.74
5-10-1	13	0.73	5.03
5-15-1	9	0.71	4.33
5-20-1	9	0.72	5.79

The reductions in the normalized mean square error during the training phase for some of structures are shown in Figure-5.

In most cases, the total solar radiation data may not always be available on long term basis, especially if there are a couple of missing data in the database, attempt was made to trained the network structures with limited solar radiation data. For this purpose, only one year (2001) data out of the 22 years solar radiation data was used in training the networks and the prediction performances of the different structures were also tested based on the prediction for the year 2007.

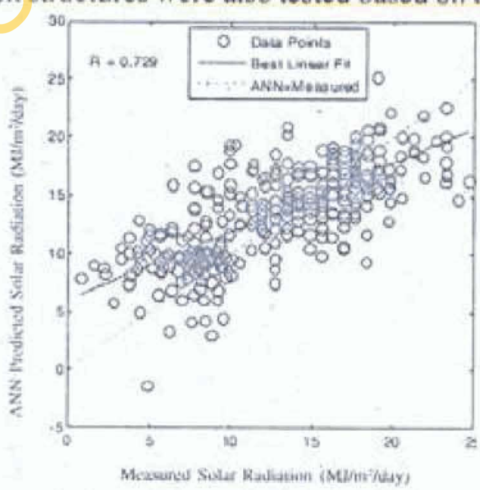

Figure-3 : Comparison of ANN prediction and actual solar radiation values for year 2007 using network structure 5: 5-10-1 (trained with 22 years data)

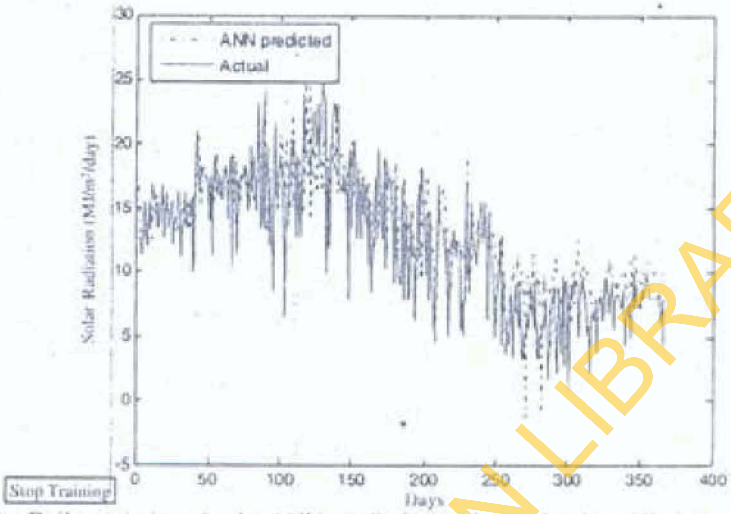

Figure-4 : Daily variations in the ANN prediction and actual solar radiation values for year 2007 using network structure 5: 5-10-1 (trained with 22 years data)

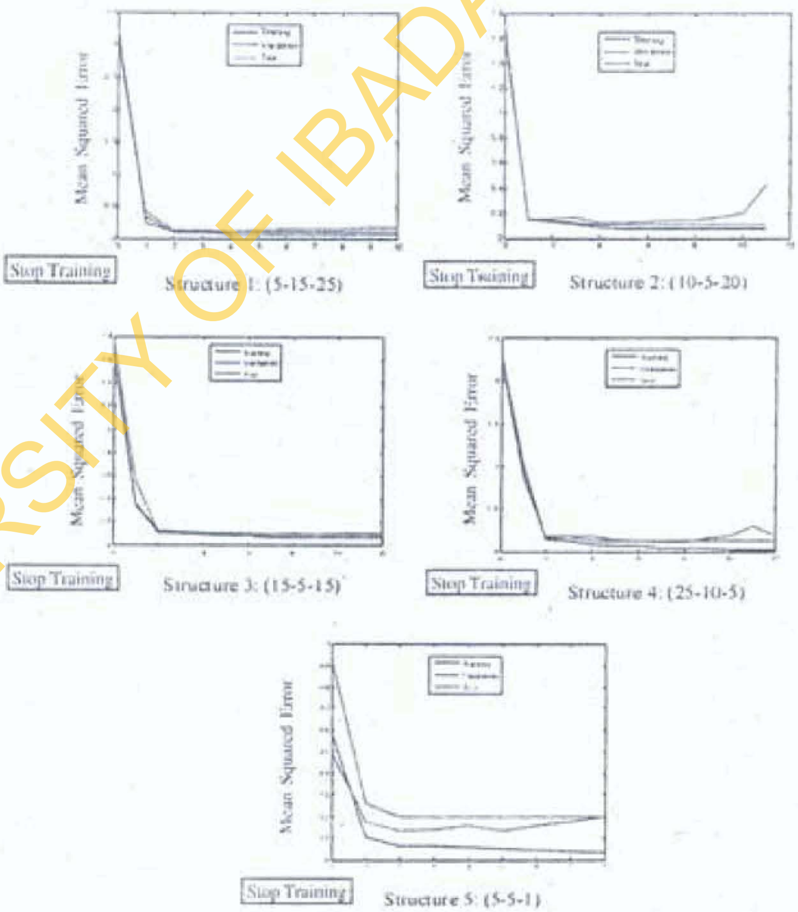

Figure-5 : Reduction in MSE during the training process for different network structures trained with 22 years data.

The performance parameters of the different structures trained with only one year data are shown in Table 3. As can be seen, that structure 5 (with 5-20-1 architecture) with R-value of 0.79 gave the best result compared to other networks investigated. Figure 6 shows a comparison between ANN predicted and actual solar radiation values for 2007 using structure 5 (with 5-20-1 architecture), while the daily variations in the predicted and actual values are shown in Figure 7. It can be seen (Figures 7), that there is a close agreement between the predicted and the actual values. The network trained with only one year data gave a better performance (R-value = 0.79) compared to that trained with 22 years data (R-value = 0.73).

Table-3 : Performance parameters of the different network structures trained with one year solar radiation data for prediction of 2007 data

Structure	No. of iteration	Correlation of coefficient (*R*-value)	MSE
Structure 1 5-15-25	6	0.43	45.54
Structure 2 10-15-20	7	0.46	66.44
Structure 3 15-15-15	8	0.39	42.62
Structure 4 25-10-5	11	0.40	21.53
Structure 5 5-5-1	9	0.79	28.38

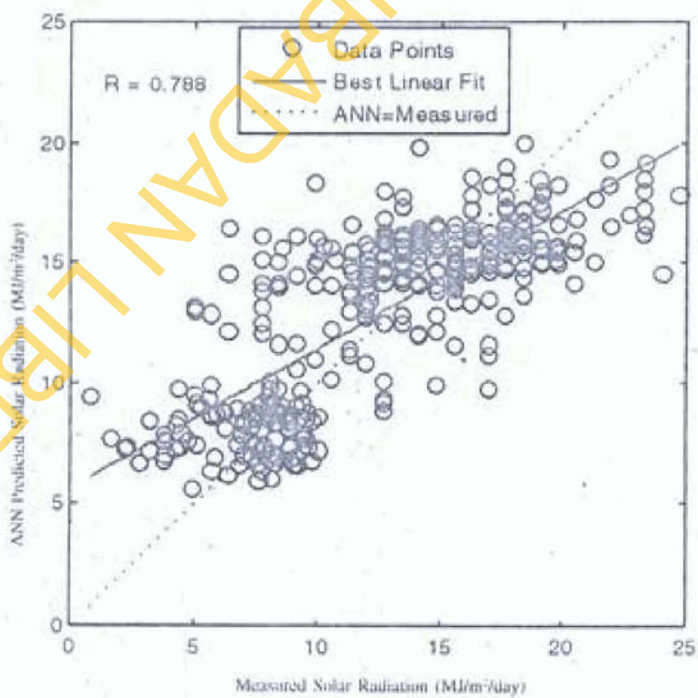

Figure-6 : Comparison of ANN predicted and actual solar radiation values for 2007 (trained with 1 year data) using structure 5 (5-5-1)

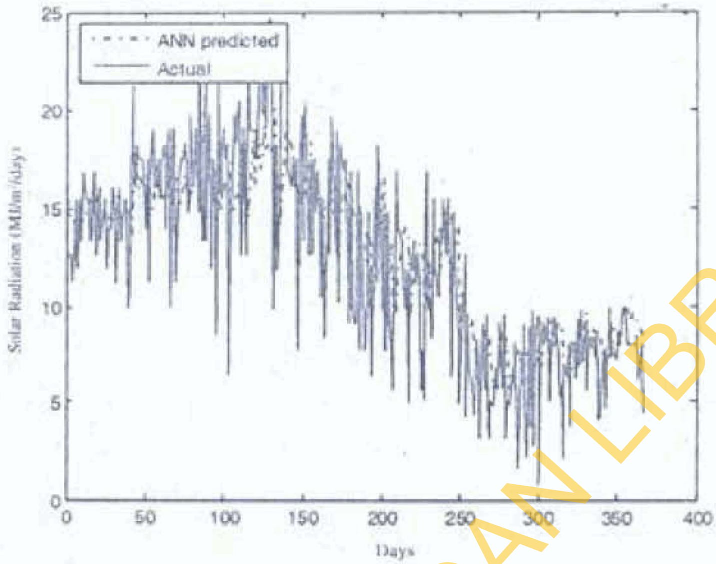

Figure-7 : Daily variations in the ANN predicted and actual solar radiation values for 2007 (trained with 1 year data) using structure 5 (5-5-1)

4 CONCLUSIONS

In this paper, a suitable method for forecasting missing daily global solar radiation data using an artificial neural network is described. The prediction of future sequences of daily solar radiation is done in a simple way. This design can predict future daily solar radiation values of 1 day based on the values of the preceding 5 days. This method is suitable for filling missing data of daily solar radiation. The validation of the model was performed with previous data, which the model has not seen before. The best correlation coefficient between the mode predictions and actual solar radiation values of 0.73 and 0.79 were obtained for structure 5 when trained with 22 years and one year data, respectively. These accuracies are within the acceptable level used by design engineers. Although these results were obtained for solar radiation of a Nigerian station, the methodology can be applied to any meteorological data and for any geographical area.

REFERENCES

1. Acock, M. C. and Y. A. Pachepsky (2000) : Estimating missing weather data for agricultural simulations using group method of data handling. *Journal of Applied Meteorology*, 39(7), pp. 1176–1184
2. Aguiar, J. M. Collares-Perrira and S. P. Conde (1988) : Simple procedure for generating of daily radiation values using a library of Markov Transition Matrices. *Solar Energy*, 49, pp. 229-279.
3. Aguiar, R. and M. Collares-Pereira (1992) : TAG : A time dependent auto regressive gaussian model for generating synthetic hourly radiation. *Solar Energy*, 49, pp. 167-74.

4. Al-Alami, S. M. and H. A. Al-Hinai (1998) : An ANN-based approach for predicting global radiation in locations with no direct measurement instrumentation. *Renewable Energy,* 14(1-4), pp. 199-204.

5. Ashraf, M., J. C. Loftis and K. G. Hubbard (1997) : Application of geostatistics to evaluate partial weather station networks. *Agric. Forest. Metoerol.,* 84, pp. 255–264.

6. Blackie, J. R. and T. K. M. Simpson (1993) : Climatic variability within the Balquhidder catchments and its effect on Penman potential evaporation. *J. Hydrol.,* 145, pp. 371–387.

7. Demuth, H. and M. Baale (2000) : Neural Network Toolbox Manual. *Mathworks Inc.* USA. 850pp.

8. Elizondo, D., G. Hoogenboom, R. Mc Clendon (1996) : Development of Neural Network to predict daily solar radiation. *Agric. Forest Meteorol.*; 71, pp. 115-32.

9. Fadare, D.A. (2009) : Modelling of solar energy potential in Nigeria using an artificial neural network model. *Applied Energy* (in press) doi :10.1016/j.apenergy.2008.12.005.

10. Greece, W. H. (2000) : *Instructional state datasets for Econometrics,* Walking Paper, Boston College, Department of Economics.

11. Kalogirou, S.A. (2001) : Artificial Neural Networks in renewable energy systems applications: a review. *Sustainable Energy Review,* 5, pp. 373-401.

12. Kohn, R. and C. F. Ansley (1986) : Fast filtering for seasonal moving average models. *Oxford Journals, Biometrika,* 73(2), pp. 522-524.

13. Little, R. J. A. and D. B. Rubin (1987) : *Statistical analysis with missing data,* John Wisely and Sons, New York, 46pp.

14. Maafi, A. and A. Adana (1989) : A two state Markovian model of global irradiation suitable for photovoltaic conversion. *Solar Wind Technol.,* 6, pp. 247-52.

15. Mellit, A., M. Benghanen and S.A. Kalogirou (2006) : An adaptive wavelet network model for forecasting daily total solar radiation. *Applied Energy* 83, pp. 705-722.

16. Mora, L. L. L. and M. Sidrash-de-Cardona (1998) : Multiplicative ARMA models to generate hourly series of global irradiation. *Solar Energy* 63, pp. 283-91.

17. Mott, P., T. W. Sammis and G. M. Southward (1994) : Climate data estimation using climate information from surrounding climate stations. *Appl. Eng. Agric.,* 10, pp. 41–44.

18. Pourhamadi, M. (1989) : Estimation and interpolation of missing values of a stationary timeseries. *Journal of Time Series Analysis,* 10, pp. 149-169.

19. Shafer, J. L. (1997) : *Analysis of incomplete multivariate data,* Chapman and Hall, London 430pp.

20. Stockinger, N. and R. Dutter (1987) : Robust time series analysis: a survey. *Kybernrtics,* 23, pp. 1-91.